Recommendation System using machine learning and movie information

2013104063 김재현

1. Background

Most people get information about movies on the Internet. People evaluate about the movie they saw, and movie site recommends movies that the customer likes based on their grade. One of the company that makes this system is Netflix. Netflix designs algorithms not only for elements such as popular personalization but also for diversity. It does not only stay in the contents that the user mainly watches but also continues to uncover the content that the user can watch and enjoy. This company also consider about when to show recommended movies for user. If they recommend them too quickly, you may be confused, and if they recommend it too slowly, you might think that you think there is nothing to see. They also care about the language of user and the culture of each country. We can experience this system named ‘Watcha’ in Korea. I want to make the most efficient movie recommendation system using the information available I can get through this project.

1. Objectives

I want to make movie recommendation system using machine learning. This system would be a system that combines the types of movies and keywords of film synopsis for movie recommendation, rather than just reflecting the ratings. For this system, I will find the most efficient method and algorithm through testing various methods.

1. Dataset Explanation

Dataset consists of the following files:[[1]](#endnote-1)

* movies\_metadata.csv: The main Movies Metadata file. Contains information on 45,000 movies featured in the Full MovieLens dataset. Features include posters, backdrops, budget, revenue, release dates, languages, production countries and companies.
* keywords.csv: Contains the movie plot keywords for our MovieLens movies. Available in the form of a stringified JSON Object.
* credits.csv: Consists of Cast and Crew Information for all our movies. Available in the form of a stringified JSON Object.
* links.csv: The file that contains the TMDB and IMDB IDs of all the movies featured in the Full MovieLens dataset.
* links\_small.csv: Contains the TMDB and IMDB IDs of a small subset of 9,000 movies of the Full Dataset.
* ratings\_small.csv: The subset of 100,000 ratings from 700 users on 9,000 movies.

1. 관련 연구

Matrix factorization methods

Some of the most successful realizations of latent factor models are based on matrix factorization. In its basic form, matrix factorization characterizes both items and users by vectors of factors inferred from item rating patterns. High correspondence between item and user factors leads to a recommendation. These methods have become popular in recent years by combining good scalability with predictive accuracy. In addition, they offer much flexibility for modeling various real-life situations. Recommender systems rely on different types of input data, which are often placed in a matrix with one dimension representing users and the other dimension representing items of interest. The most convenient data is high-quality explicit feedback, which includes explicit input by users regarding their interest in products.[[2]](#endnote-2)

1. 참고문헌

https://www.kaggle.com/rounakbanik/the-movies-dataset

Koren, Yehuda, Robert Bell, and Chris Volinsky. “Matrix factorization techniques for recommender systems.”, 2009

1. https://www.kaggle.com/rounakbanik/the-movies-dataset [↑](#endnote-ref-1)
2. Koren, Yehuda, Robert Bell, and Chris Volinsky. “Matrix factorization techniques for recommender systems.”, 2009 [↑](#endnote-ref-2)